Capstone Proposal

**NHL Player Analysis**

**Business Understanding**

NHL owners need to evaluate player statistics when signing or re-signing existing players. Through machine learning we can develop a system to predict player contract prices allowing team owners more insight through contract negotiations. This projects target audience will be NHL management or any person in within a team’s organization that has contract negotiation powers. This model will impact a team by giving them the capable resources to identify player contract prices.

**Data Understanding**

The data will be collected through the NHL’s API and scraped from Spotrac’s website. Spotrac has extensive statistical information for any professional sport. A few features in the dataset will include player name, position, jersey number, salary, height, & weight.

**Data Preparation**

The data will be stored and cleaned within a pandas dataframe. Preprocessing will include converting JSON files to dataframes and merging tables based on common columns. I expect the dataframe to be ~950 rows of active players. At this time, I do not have data on inactive players.

**Modeling**

Since this will be a regression problem the project will be using Linear Regression, Random Forest, SVM, XGBoost, and Neural Networks. Our target variable will be the players salary. To start, we will use a Linear Regression model as our baseline and move forward using pipelines to find our best model.

**Evaluation**

Our evaluation will be based on RMSE. I believe this task can be completed within a week. My stretch goals would be to incorporate a UI allowing end users to filter stats based on player. Also, incorporating some form of time series forecast would a huge stretch goal.

**Deployment**

Deployment will consist of a user interface that will be filtered based on player.

**Tools/Methodologies**

We will be using all of the data science tools a few are:

* Sklearn
* Pandas
* XGBoost
* Matplotlib

The analysis will be completed on my local machine storing the data locally and on github.

**Business Problem:**

NHL Teams need leverage when negotiating contract prices. Through the use of machine learning we can review all salaries within the NHL to predict a player’s salary and projected salaries for a better cap space understanding. Also, a team needs an understanding of how this player will affect their overall performance. I.E. where do they score the most on the ice? Will this compliment their line mates? Through visual analysis we can build a team roster that has strong offensive capabilities.

**Business Insights:**

When does a player’s value drop due to age?

Injuries? Are more players prone to injuries vs others? Can we see a trend in value?

Do free agents get higher contracts vs resigning with their existing team?

What makes a player valuable?

**Stakeholders:**

Players

Player’s managers

coaches

Nhl team management

Owners

**Modelling Techniques:**

Baseline: Linear Regression

Ridge/Lasso Regression

Decision Tree Regressor

SVM

Neural Network

Prophet

**Deployment:**

The project will be deployed using Tableau. I envision an interactive dashboard that allows end users to filter players based on contract price that fits within their salary cap, projected salaries within contract duration, and a page detailing the players offensive capabilities. I.E. where the best place for them to score on the ice? Alternatively, we can look at where goalies get scored on the most.